基于注意力机制的音乐深度推荐算法 *

张全贵，张新新，李志强

(辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125105)

摘　要：在海量音乐中，如何根据用户的历史收听记录分析用户需求以实现歌曲推荐是音乐推荐领域具有挑战性课题之一。现有的音乐推荐方法仅简单将用户听过的所有音乐均作为音乐推荐的上下文，导致不同类型音乐学习到的上下文权重分配相同，其严重影响了音乐推荐精度。针对此问题，提出了一种基于注意力机制的音乐深度推荐方法，针对不同用户的历史收听音乐动态分配不同的注意力，即学习出不同的上下文权重，使推荐结果更符合用户的实际偏好。通过在公开音乐数据集 Million Song Dateset 上的测试，所提方法的推荐准确率有很大的提升。

关键词：深度学习；注意力机制；音乐推荐

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2018.01.0075

Music recommendation algorithm based on attention mechanism

Zhang Quangui, Zhang Xinxin, Li Zhiqiang

(*School Electronics & Information Engineering Liaoning Technical University*, *Huludao Liaoning* 125105, *China*)

Abstract: In the mass music, how to analyze the user's needs according to the user's history listening record to implement song recommendation is one of the challenging topics in the music recommendation field. The existing music recommendation method simply uses all the music the user has heard as the context of the music recommendation, which results in the same weight distribution of contexts learned by different types of music, which seriously affects the accuracy of the music recommendation. In response to this problem, this paper proposed a music recommendation method based on attention mechanism, which dynamically allocated different attentions to different users' historical listening music, that was learns different contextual weights so as to make the recommendation result more in line with the user's actual preference . And through the test on public music dataset named Million Song Dateset, the recommended accuracy of the method proposed in this article has greatly improved.

Key words: deep learning; attention mechanism; music recommendation

0　引言

音乐，是我们日常生活中不可或缺的一部分，每个人都有自己独特的音乐偏好，由于受个人性格、成长环境及所受教育程度等不同因素的影响，每个用户可能喜欢多种类型的音乐，所以用户的音乐偏好可以从他的历史收听记录中加以准确分析。在现实世界中，用户要听的下一首歌曲可能只与历史播放记录中的部分歌曲相关，而传统的音乐推荐方法往往是把用户收听的历史记录中所有音乐均作为上下文信息。显然，与用户实际偏好是不相符的。

近些年，深度学习(deep learning，DL)在许多应用领域已经取得了巨大的成功，如计算机视觉、机器翻译等。目前，深度学习技术也广泛应用在推荐系统领域[6,7]，它可以有效地捕获非线性和非平凡的用户项目关系，并进行更复杂抽象的高层数据表示[5]。此外它还可以从原始数据中学习更复杂的用户偏好，从而增强推荐效果。基于深度学习的推荐方法近年来得到了很大关注，推荐系统领域的国际会议 RecSys 自 2016 年开始组织推荐系统的深度学习研讨会，旨在推动和鼓励基于深度学习的推荐系统的研究与应用。

在音乐推荐(music recommendation，MR)领域，现有的基于深度学习的音乐推荐方法，使用音乐本身的音频信息[8]、专辑信息、风格、艺术家、歌词和用户对音乐的评价、关键词和收听历史记录[1]等。将音乐的音频信号输入深度卷积神经网络中学习来预测用户的音乐偏好的研究方法[8]证明了深度学习非常适合音乐推荐，而且深度卷积神经网络明显优于传统方法。尽管音乐内容对于人们的音乐收听偏好是非常重要的，但由于

收稿日期：2018-01-30；**修回日期：**2018-03-28　　**基金项目：**国家留学基金资助项目（留金法［2015］5104）

作者简介：张全贵（1978-），男，河北秦皇岛人，副教授，博士，主要研究方向为计算机视觉、推荐系统、机器学习、数据挖掘等（zhqgui@126.com）；张新新（1994-），女，辽宁朝阳人，硕士，主要研究方向为机器学习、推荐系统、数据挖掘等；李志强（1993-），男，辽宁锦州人，硕士，主要研究方向为机器学习、计算机视觉、推荐系统等.

事实上音乐特性与相应的音频信号之间存在较大的语义差异，也忽略了不同用户之间兴趣相似的情况。而且对音频内容的分析，需要将音频信息转换为数字信息 这也会导致计算量增大、响应时间延长等问题，不能在大规模系统中实现快速预测，因此基于音乐内容的推荐方法并不能很好地预测到用户对下一首音乐的偏好。但是传统基于用户历史行为的音乐推荐中，针对不同的用户及要预测的音乐给用户的历史音乐列表上下文都学习出的相同权重[9]，然而这样忽略了用户音乐偏好的多样性，即用户可能会喜欢不同类型的音乐，因此下一首将要收听的音乐可能只与之前听过的音乐中的一部分相关。而注意力机制与深度学习技术的融合对音乐推荐研究有重大改进。

注意力机制最早是在视觉图像领域提出的[10,11]，真正得到关注的是 Google mind 团队提出的将 AM 融合进 RNN 模型中进行图像分类[12]。近几年也被应用于文本翻译[13]、自然语言处理[14]等领域。图 1 为 AM 在神经网络机器翻译中实现原理[14]。在推荐系统领域，注意力机制表现也是可贺的，基于注意力机制的 CNN 能够在输入中捕获到最有用的元素信息[13]；文献[16]提出一种基于注意力机制的融合性深度神经网络，将 Twitter 与用户兴趣之间的用户、作者、用户兴趣相似性信息融合作为学习属性，来预测用户的转发行为。至今将注意力机制与深度学习模型的结合方式应用在音乐推荐领域的研究甚少，文献[15]提出了一种根据意图重要性将注意分数分配给不同的单词位置的注意力机制网络模型，其中序列的矢量输出由每个单词根据其注意分数的 RNN 的隐藏状态的加权和计算用于提高复调录音的准确性。针对上述问题，受到计算机视觉及自然语言处理中注意力机制的启发，提出一种基于 attention 机制的音乐深度推荐算法，通过对用户历史播放记录加入注意力机制分析，实现对历史记录中的音乐分配动态权重，辅以推荐建模。

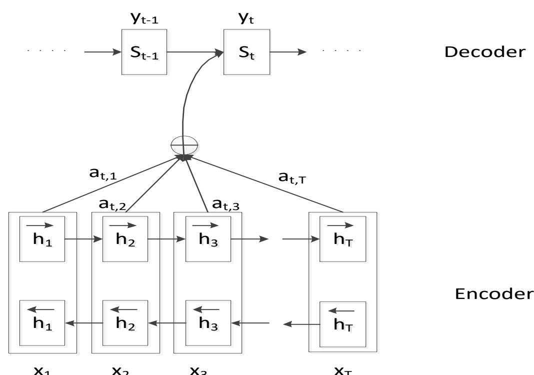


图 1　神经网络机器翻译中 Attention 机制设计部分

1　AMNN 推荐算法

基于注意力机制的音乐深度推荐算法（简称 AMNN）是在深度学习 DNN 算法的基础上而提出的，模型架构如图 2 所示。用户、用户历史收听记录及待预测音乐作为深度神经网络的输入，通过嵌入层得到它们的特征表示，借鉴 NeuMF[17]的思想，通过多层网络学习用户与音乐及音乐与音乐之间的个性化关系。为了学习到动态的用户偏好音乐权重，Attention 机制聚合了用户历史收听记录的音乐信息的表示来表征用户的音乐偏好权重，从而判断新输入到模型中的音乐是否符合用户喜好。

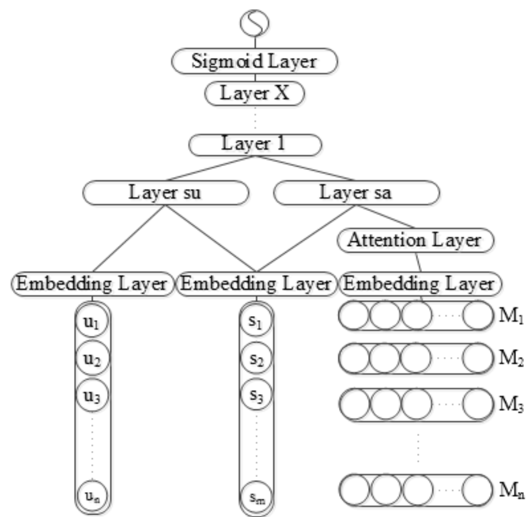


图 2　AMNN 模型架构

问题定义：假设用户集合 $U=\{u_1,u_2,\ldots\ldots,u_n\}$，音乐集合 $S=\{s_1,s_2,\ldots\ldots,s_m\}$。而对每个用户 u，都有对应的历史音乐收听列表 $M_u=\{S^1{}_u,S^2{}_u,\ldots\ldots,S^k{}_u\}$，其中 $S^k{}_u$ 是用户 u 收听的第 K 首音乐歌曲。本文要做的预测问题是给出用户的历史收听列表 M_u，为每个用户 u 预测适合他的下一首音乐 S_u。

用户 ID、音乐 ID 和该用户的历史收听列表中 K 首音乐 ID 作为模型的起始输入，将输入列表转换为分布式向量表示，其中每个用户被表示为 $u \in R^{d_u}$，每首音乐被表示为 $s \in R^{d_i}$，而用户的收听列表为 $M_u=\{S^1{}_u,S^2{}_u,\ldots\ldots,S^k{}_u\}$，$s_u^i \in R_u^{d_i}$。通过将所有的用户、音乐和用户音乐收听矢量列表分别放在一起，可以得到用户矩阵 $U \in R^{d_u*|U|}$、音乐矩阵 $S \in R^{d_i*|S|}$ 以及用户的音乐历史收听列表矩阵 $M_U = \{S_u^1,S_u^2,\ldots\ldots,S_u^k\}, S_u^k \in R_u^{d_i^k*|S|}$。并将得到的输出矩阵 U 和 S 作为非线性隐藏层的输入，同时 M_u 中每首音乐作为 Attention 机制的输入。

Attention 机制可以学习历史收听记录中每首音乐对预测下一首音乐所起的不同作用，即注意力。如式(1)所示，用户 u 对第 i 首音乐的偏好函数 a_u 为

$$a_u^i = \tanh(w_i S_u^i + b_i) \tag{1}$$

其中：$S_u{}^i$ 表示用户 u 的历史收听列表中的第 i 首音乐；w_i 和 b_i 是可学习的参数，分别用来表示该首历史音乐对待预测的下一首音乐所起作用的权重和偏置。通过 tanh 函数得到每首音乐对待预测的下一首音乐的非线性关系，再利用 softmax 函数对其进行归一化，使得所有历史音乐对待预测的下一首音乐所起的总体作用限制在 1，从而得到用户 u 对音乐 i 的偏好权重 α^i。

$$\alpha_u^i = \frac{\exp(a_u^{i\,T} u_w)}{\sum_n \exp(a_u^{i\,T} u_w)} \tag{2}$$

其中：u_w 为预先设定的权重。最后将用户 u 的整个音乐列表的偏好权重求和得到权重向量φ。

$$\varphi = \sum_k s_u^k \alpha_u^k \quad (3)$$

在非线性隐藏层使用 Relu 作为激活函数，首先用户特征向量和待预测的音乐特征向量以及音乐特征向量和通过 Attention 机制的用户音乐偏好权重向量分别进行学习，然后并入到多层深度网络中。定义其预测模型为 $y_{u,s} = f([su, sa]u, s)$，前向传播过程为 $v_1 = \phi_1([su, sa])$，su、sa 分别表示用户 u 和待预测音乐 s 以及待预测音乐 s 和通过 Attention 机制的用户音乐偏好权重φ的输出结果，则

$$\phi_2(v_1) = \sigma(w_2 v_1 + b_2) \quad (4)$$

$$\phi_h(v_{h-1}) = \sigma(w_h v_{h-1} + b_h) \quad (5)$$

其中：h 是层的深度；σ 是激活函数。经过 X 层网络学习后，最终的映射函数为

$$y_{u,s} = \sigma(W_x(\Gamma(w_2[U,S,\varphi] + b_2)) + b_x) \quad (6)$$

其中：Γ 函数代表 *X-1* 层激活函数嵌套。

在最后的输出层，通过 sigmoid 激活函数计算出的带预测的音乐特征向量是符合用户喜好的下一首音乐的可能性，预测值 $y_{u,s} = [0,1]$，定义似然函数为

$$P(O, O^-|U, S, \varphi) = \prod_{(u.i)\epsilon O} y_{u,s} \prod_{(u,i)\epsilon O^-}(1 - y_{u,s}) \quad (7)$$

上式取似然负对数作为损失函数可得的交叉熵损失函数为

$$L = -\sum_{(u,i)\in O \cup O^-}[y\log(y_{u,s}) + (1-y)\log(1-y_{u,s})] \quad (8)$$

2　实验设计和效果分析

针对本文提出的基于深度学习和注意力机制的音乐推荐方法的优越性验证，在 Cluster 的一个 node 上实现了所提出的算法。实验环境如表 1 所示。将所提出的 AMNN 算法与 itemKNN、NeuMF 和 NNrec 等推荐算法的等推荐算法在同实验设置下从不同角度进行了实验对比，最后对数据结果进行比对分析和原因预测。

2.1　数据集

本文使用的数据集是公开音乐数据集 Million Song Dataset(MSD)，它包含来自 SecondHandSongs dataset、musiXmatch dataset、Last.fm dataset、Taste Profile subset、thisismyjam-to-MSD mapping、tagtraum genre annotations 和 Top MAGD dataset 七个知名音乐社区的数据。而本文主要使用 MSD 的核心数据，即 The Echo Nest 发布的 Taste Profile 子集，它由三元组（用户 ID、音乐 ID、用户历史收听记录）组成，其中包含了 110 000 名用户、386 213 首歌曲和 1 450 933 条用户历史收听记录。为了提高模型的训练速度，在数据进行预处理。首先对音乐历史收听记录不少于 20 首的用户进行采样，从而得到 19 502 名用户、115 082 首音乐和 390 040 条用户历史收听记录；然后实验过程采取 5 折交叉验证方案进行训练和测试，即随机将数据集分成 5 份，并依次迭代地将其中 4 份作为训练集，其余的作为测试集。

2.2　参数设置

在 AMNN 模型中，隐藏层的激活函数选用 relu 函数，它收敛速度快，易于稀疏激活。模型最终的预测结果为二进制分类，所以使用 sigmoid 函数作为最后输出层的激活函数。训练过程中利用 Adam 优化方法优化模型，因 Adam 的优点是能在经过偏置校正后，使每一次迭代学习率都有一个确定范围，这样得到的参数比较平稳，实验中学习率设置为 0.003。最后为防止模型在计算过程中出现过拟合现象，利用 L2 正则化方法对网络参数进行约束，并加入了 dropout 策略，其中取 dropout rate=0.5。

表 1　实验环境

实验环境	环境配置
操作系统	Linux
主频	3.5 GHz
内存	96.00 GB
编程语言	Python 3.5
深度学习框架	Tensorflow 1.2.0、Keras 2.0.5

2.3　实验结果度量标准

本文设计的目的是为了给用户推荐更加精确的下一首音乐，每个用户去除一首听过的音乐作为测试音乐，然后在随机抽取 99 首没有听过的音乐作为负例音乐，这样每个用户共有 100 首音乐作为测试数据，从而组成一个待推荐的音乐列表，其中有一个正例和 99 个负例。

推荐性能由命中率（HR）和归一化折扣累积增益（NDCG）来衡量。本文将这两个指标的排名列表截断 *K* 均为 10。因此 HR 直观地衡量测试歌曲是否存在于前 10 项列表中，而 NDCG 是用来衡量排序质量的指标，通过将较高分数指定为顶级排名来计算命中的位置。命中率(HR)和归一化折扣累积增益(NDCG)的定义如下：

命中率(HR)为

$$HR@K = \frac{Number \quad of \quad Hits@K}{|GT|} \quad (9)$$

归一化折扣累积增益(NDCG)为

$$NDCG@K = Z_K \sum_{i=1}^{K} \frac{2^{r_i} - 1}{\log_2(i+1)} \quad (10)$$

其中：GT 表示测试歌曲与算法产生的歌曲顺序列表的集合；Z_K 是正则化，以确保排名列表中存在一个值 1；r_i 是第 i 个歌曲的预测相关性。在评估过程中使用简单的二进制表示：如果歌曲在测试集 r_i=1，否则为 0。

2.4　对比实验

-ItemKNN[18]：基于项目的协同过滤标准方法，发现音乐之间的相似度，推荐类似的音乐类型；

-NeuMF：基于神经网络协同过滤方法，对比 Attention 机制对推荐效果的影响；

-NNrec[4]：受到在 NLP 领域的基于 NN 的概率语言模型的

启发，最早被提出作为下一个 basket 的推荐算法，并且在两个零售数据集上表现出色。

在 AMNN 实验设计中，本文选取用户历史收听记录中次数的前 *K* 个作为预测用户音乐偏好的注意力分析列表，因为用户对音乐有很强的动态性，随着环境、心情、身体状态等变化，但还是他的音乐习惯是较为固定的，所以以他过去收听的歌曲次数作为注意力分析列表是可以的。

2.5　效果分析

为获取更加鲜明的实验效果对比，在 NNrec 和 AMNN 实验中选取用户历史收听记录的前 10 首音乐。通过推荐性能 HR(图 3(a))和 NDCG(图 3(b))的效果图中，可以看到随着 *K* 的增加，itemKNN 始终最差，与其作比较，NeuMF 可以达到 0.657 8 和 0.453 7，而 NNrec 要比它提高 3.4%和 2.3%，但是 AMNN 要比 NNrec 还提高 1.6%和 4.2%。因此验证了 AMNN 优于现有方法，可以提高推荐质量。

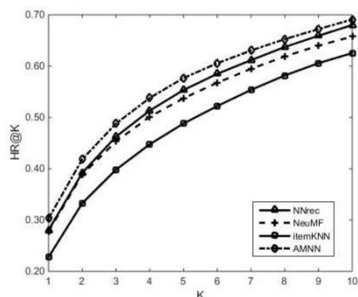


图 3(a)

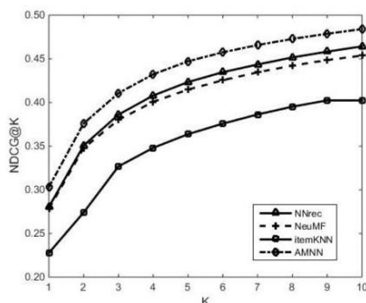


图 3(b)

图 3　TOP-K 值对模型推荐性能的影响

本文也考虑到历史收听列表中的音乐数量可能是影响实验效果的一个重要因素。接下来，对融入 AM 的 AMNN 与未融入 AM 的 DNN 作对比，其中 K 取为 10，最终得到了证明。其效果见图 4。对于本文使用的数据集来说，获取的用户历史音乐收听记录在 15 首左右时就可以更好地学习到用户的个人音乐偏好，而未融入 AM 的 DNN 实验只有获取越多的用户历史音乐收听记录才能更好地学习，验证了 AMNN 可以在获取用户的部分收听记录情况下就能学习到用户的音乐偏好，从而降低了计算复杂度。

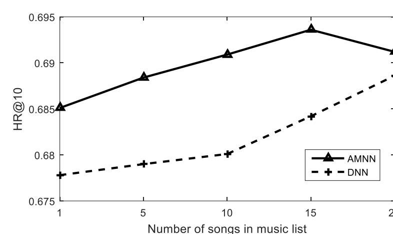


图 4(a)

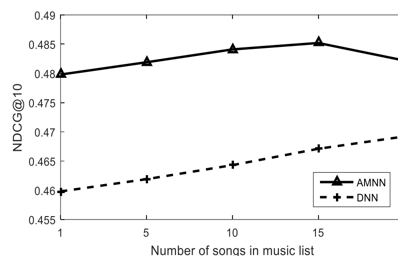


图 4(b)

图 4　历史收听列表中的音乐数量对模型推荐性能的影响

3　结束语

本文提出了一种基于注意力机制的音乐深度推荐方法，在基于深度学习的音乐推荐算法中加入能够为用户分配不同音乐偏好权重的注意力机制来实现更准确的推荐。通过在 MSD 数据集上实验表明，Attention 机制对提高推荐精度有一定的效果，并在一定程度上使对用户的兴趣偏好的分析更有可解释性。在未来工作中，将考虑多种注意力机制（如局部注意力机制）对用户偏好获取的影响，并融入到基于深度学习的音乐推荐算法，进一步提高推荐质量。

参考文献：

[1] 陈雅茜. 公共环境下的混合型音乐推荐系统的关键技术研究 [J]. 计算机应用研究, 2012, 29 (11): 4250-4253. (Chen Yaqian. Research on key technologies of hybrid music recommendation system in public environment [J]. Application Research of Computers, 2012, 29 (11): 4250-4253.)

[2] 滕少华, 郑明, 刘冬宁. 面向音乐推荐的全变差图非负矩阵分解方法 [J]. 计算机应用研究, 2018, 35 (4) . (Teng Shaohua, Zheng Ming, Liu Dongning. Non-negative matrix factorization method for total variation maps for music recommendation [J]. Journal of Computer Applications, 2018, 35 (4) .)

[3] Wan Shengxian, Lan Yanyan, Wang Pengfei, *et al.* Next basket recommendation with neural networks [C]// Proc of RecSys Posters. 2015: 1-4.

[4] Hsu K C, Chou S Y, Yang Y H, *et al.* Neural network based next-song recommendation [J]. Computer Science. 2016: 35-40.

[5] Zhang Shuai, Yao Lina, Sun Aixin. Deep learning based recommender system: a survey and new perspectives [Z]. 2017: 182-187.

[6] Covington P, Adams J, Sargin E. Deep neural networks for youtube recommendations [C]// Proc of the 10th ACM Conference on Recommender

Systems. 2016: 191-198.

[7] Okura S, Tagami Y, Ono S, *et al.* Embedding-based news recomm-endation for millions of users [C]// Proc of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017: 1933-1942.

[8] Van Den Oord A, Dieleman S, Schrauwen B. Deep content-based music recommendation [C]// Advances in Neural Information Processing Systems. 2013: 2643-2651.

[9] Liang D, Zhan M, Ellis D P W. Content-aware collaborative music recommendation using pre-trained neural networks [C]// Proc of ISMIR. 2015: 295-301.

[10] Larochelle H, Hinton G E. Learning to combine foveal glimpses with a third-order Boltzmann machine [C]// Advances in Neural Information Processing Systems. 2010: 1243-1251.

[11] Denil M, Bazzani L, Larochelle H, *et al.* Learning where to attend with deep architectures for image tracking [J]. Neural Computation, 2012, 24 (8): 2151-2184.

[12] Mnih V, Heess N, Graves A. Recurrent models of visual attention [C]// Advances in Neural Information Processing Systems. 2014: 2204-2212.

[13] Luong M T, Pham H, Manning C D. Effective approaches to atten-tion-based neural machine translation [C]// Proc of EMNLP. 2015: 34-41.

[14] Bahdanau D, Cho K, Bengio Y. Neural Machine translation by jointly learning to align and translate [J]. Computer Science, 2014: 394-402.

[15] Zhai Shuangfei, Chang Kenghao, Zhang Ruofei, *et al.* Deepintent: learning attentions for online advertising with recurrent neural networks [C]// Proc of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 1295-1304.

[16] Zhang Qi, GongYeyun, Wu Jindou, *et al.* Retweet prediction with attention based deep neural network [C]// Proc of the 25th ACM International on Conference on Information and Knowledge Management. 2016: 75-84.

[17] Vall A, Eghbal-zadeh H, Dorfer M, *et al.* Music playlist continuation by learning from hand-curated examples and song features [C]// Proc of the 2nd Workshop on Deep Learning for Recommender Systems. 2017: 46-54

[18] Sarwar B, Karypis G, Konstan J, *et al.* Item-based collaborative filtering recommendation algorithms [C]// Proc of International Conference on World Wide Web. 2001: 285-295.

[19] Vinyals O, Kaiser L, Koo T, *et al.* Grammar as a foreign language [J]. Eprint Arxiv, 2014: 2773-2781.

[20] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J]. Computer Science, 2014: 738-745.